A Low-Complexity Mosaicing Algorithm for Stock Assessment of Seabed-Burrowing Species

David Corrigan, Ken Sooknanan, Jennifer Doyle, Colm Lordan, and Anil Kokaram

***Abstract*—This paper proposes an algorithm for mosaicing videos generated during stock assessment of seabed-burrowing species. In these surveys, video transects of the seabed are captured and the population is estimated by counting the number of burrows in the video. The mosaicing algorithm is designed to process a large amount of video data and summarize the relevant features for the survey in a single image. Hence, the algorithm is designed to be computationally inexpensive while maintaining a high degree of robustness. We adopt a registration algorithm that employs a simple translational motion model and generates a mapping to the mosaic coordinate system using a concatenation of frame-by-frame homographies. A temporal smoothness prior is used in a maximum *a posteriori* homography estimation algorithm to reduce noise in the motion parameters in images with small amounts of texture detail. A multiband blending scheme renders the mosaic and is optimized for the application requirements. Tests on a large data set show that the algorithm is robust enough to allow the use of mosaics as a medium for burrow counting. This will increase the verifiability of the stock assessments as well as generate a ground truth data set for the learning of an automated burrow counting algorithm.**

***Index Terms*—Feature-based image registration, Nephrops surveys, underwater mosaicing, underwater television (UWTV).**

I. Introduction

STOCK assessment of commercially viable marine species is an essential part of fishery management and is mandated in many international jurisdictions. The methodology for these assessments varies widely and includes analysis of catch rates and sampling of landings and discards from trawls as well as sonar and video-based surveys [1]. Seabed-burrowing creatures pose a challenge for stock assessment. *Nephrops Norvegicus* (referred to from now on by its genus Nephrops) is an example of such a species that forms a commercially significant fishery. The annual catch for Nephrops is approximately 50 000 t with a first catch value of €300 million per annum [2]. Due to variations in growth and burrow emergence of Nephrops, traditional assessment techniques such as trawl catch rates and sampling are not considered to be a reliable estimator of population abundance. In recent years, marine scientists have developed a video-based population survey technique known as underwater television (UWTV) [3], [4] to overcome these limitations.

In an UWTV survey, a sledge with mounted cameras and lights is towed by a trawler, capturing a video along a transect of the seafloor. Subsequently, marine scientists review the video and manually tally the number of observed Nephrops burrows for each minute of video. The burrow density estimates at a number of sites are used to obtain an overall population estimate for each of the Nephrops fishing grounds.

Although UWTV has become a widely adopted standard for Nephrops stock assessment, there is a desire within the community to further improve counting methodologies. Manual counting is a labor-intensive task.[1] It is also a largely subjective task and can be adversely affected by the quality of underwater videos and presence of burrows of other marine species, even though standard procedures and methodologies are in place to train counters and monitor their consistency. As the information gleaned from the surveys is limited to per-minute count totals, there is no record of which objects have been counted. Furthermore, the surveys do not quantify other useful information such as burrow sizes that could be used to monitor the dynamics of the population.

In recent years, image processing and computer vision techniques have been developed to assist these surveys. The goal is to automate as much of the counting procedure as possible and supervised learning based algorithms have been developed to detect burrows automatically [6]–[8]. Content summarization in the form of mosaics of the seabed transect is another useful technique. In [7], we showed how mosaics can be used in an automated burrow detection pipeline to convert a video object detection problem into a simpler image object recognition problem. Mosaics are also a potentially useful medium for manual burrow counting as they allow for easy annotation and nonsequential counting of burrows.

In this paper, we describe a mosaicing algorithm specifically developed for UWTV surveys and is an extension of the work

Manuscript received January 27, 2017; revised August 17, 2017 and December 27, 2017; accepted February 16, 2018. This work was supported by the Science Foundation Ireland under Award SFI-PI 08/IN.1/I2112. *(Corresponding author: David Corrigan.)*

Associate Editor: J. Cobb.

D. Corrigan was with the Department of Electrical and Electronic Engineering, Trinity College Dublin, Dublin 2, Ireland. He is now with the Video Competence Centre, Huawei Irish Research Centre, Dublin 4, Ireland (e-mail: corrigad@tcd.ie).

K. Sooknanan is with the Centre for Information and Communication Technology, University of Trinidad and Tobago, Port of Spain-JDSTI, Trinidad and Tobago (e-mail: ken.sooknanan@utt.edu.tt).

J. Doyle and C. Lordan are with the Marine Institute, Rinville, Oranmore, Co., Galway H91 R673, Ireland (e-mail: jennifer.doyle@marine.ie; Colm.Lordan@marine.ie).

A. Kokaram is with the Department of Electrical and Electronic Engineering, Trinity College Dublin, Dublin 2, Ireland (e-mail: anil.kokaram@tcd.ie).

Digital Object Identifier 10.1109/JOE.2018.2808973

[1] Up to 4 000 min of video are counted each year in Ireland which has roughly 18% of the annual catch in the European Union (EU) [2], [5]. Each minute is counted by at least two marine scientists.

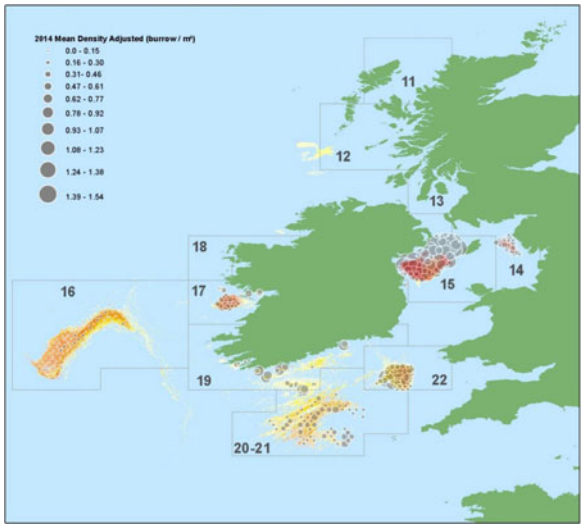

Fig. 1. FUs for Nephrops management (adapted from [10]). The figure shows a representation of the division of the nephrops fishing grounds into FUs for ICES area 7. A similar division into FUs is performed in other ICES areas with significant Nephrops fisheries. The gray circles represent the burrow density estimates for each of the stations of the 2014 UWTV surveys.

presented in [9]. Due to the large area of seabed to be processed, computational efficiency and robustness of the algorithm are key design considerations. Therefore, we employ a simpler motion model to ensure robustness and use a temporal smoothness prior on the motion parameters instead of more computationally intensive global registration approaches. The algorithm is also adapted to deal with the poor quality of underwater video and video compression. We tested the performance of the algorithm by processing an entire annual UWTV survey of all Nephrops grounds in Irish territorial waters consisting of 366 videos, each of 10 min duration. We show how these mosaics can be used to annotate Nephrops burrows and, for the first time, show how using the mosaics has allowed experts to compare their performance at a burrow-by-burrow level.

The rest of this paper is organized as follows. In Section II, we present a brief description of the video capture methodology for UWTV surveys. This is followed by a discussion on the state of the art in mosaicing including techniques specifically targeted at underwater videos (see Section III). A description of the algorithm is given in Section IV. Section V presents an analysis of the performance of the algorithm and is followed by a discussion of the potential impact of the algorithm on burrow counting in Section VI. Final remarks are given in Section VII.

II. Underwater Television Video Surveys

UWTV surveys are performed annually to estimate the Nephrops population. Nephrops fishing grounds are broken down into functional units (FUs) and separate annual assessments and catch advice are produced for each. The FUs for the International Council for the Exploration of the Sea (ICES) area 7 are shown in Fig. 1. For a given FU, a survey is performed by estimating the burrow density using a video-based methodology at a number of sampled locations spread across the known Nephrops habitat.

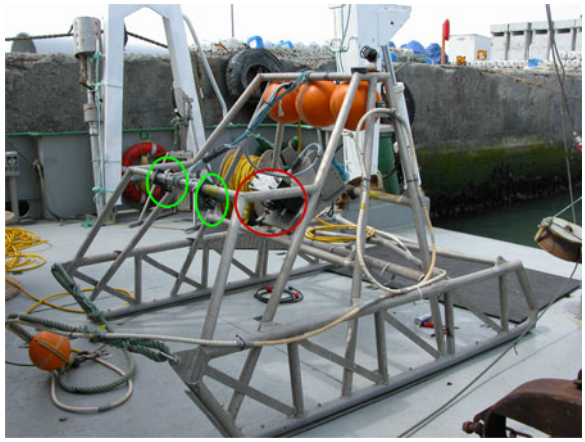

Fig. 2. Camera sledge used in UWTV surveys (adapted from [4]). The camera is highlighted by the red circle and the two artificial lights on the sledge are shown by the green circles.

At each survey location, known as a station, a camera mounted on a sledge with artificial lights (see Fig. 2) is dropped off the back of a vessel and towed in a straight line along the seafloor. Once the sledge is stable on the seafloor, a 10-min video covering approximately 200 m is captured and stored in a compressed video format [3], [4]. Video resolution and video formats vary for each national fishery management agency. Examples of frames taken from a station video are shown in Fig. 3.

Fig. 4 gives an overview of the visual characteristics of Nephrops burrows [11]. A nephrops burrow can have one or more typically crescent-shaped entrances arranged in a radial pattern around a raised center. Each 10-min station video is counted independently by two expert counters, generally when on-board the vessel. As the video is played back, the counters log observed nephrops burrows using a mechanical counter. The counters also record the number of nephrops observed within burrow entrances and on the seafloor. Only burrows which pass through the bottom of the video frame are counted. For each minute of video, the count totals are recorded along with subjective classifications of other relevant information such as visibility, sediment type, and the amount of time for which counting was possible [12], [13]. This information is then used to estimate a burrow density for the station which is in turn used to estimate the total population for each FU.

Given the economic importance of the estimated count totals, quality assurance of UWTV is a key concern. At the present time, quality control is performed by comparing intercounter variability between the two expert counters [12], [13]. Furthermore, before each counter performs the count for a particular FU, they are required to count a standard video sequence with a known count total (as recommended in [4]). This allows the counters to retrain themselves to the particular visual characteristics of burrows for that unit, which can vary in appearance due to population density, size, or the abundance of other burrowing species. However, without a record of which burrows have been counted, the effectiveness of quality control measures is limited.

The use of mosaics has the potential to improve the quality of the UWTV surveys. As annotating objects of interest in a mosaic

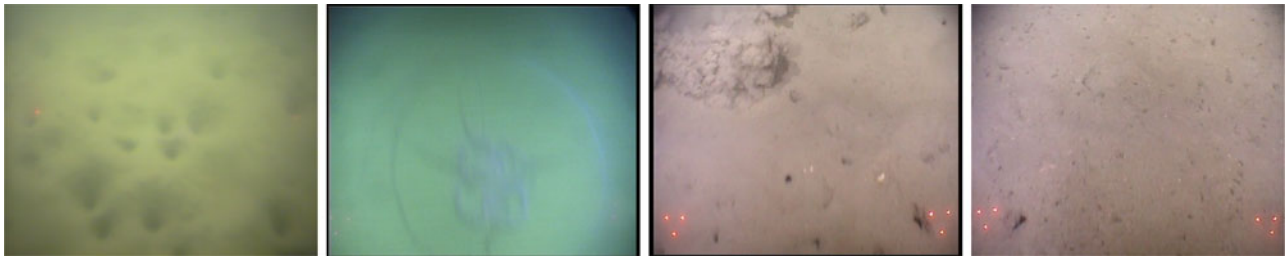

Fig. 3. Frames extracted from four stations from different FUs. From left to right: FU 15 (western Irish sea), FU 17 (Galway Bay), FU 19 (Galley Grounds), and FU 22 (Smalls fishing ground). These frames show the range of video quality and burrow distribution across the different fishing grounds. The burrow density in the western Irish Sea is much greater than that on the Porcupine Bank but the nephrops are also generally smaller in size. The sets of red dots are generated by lasers mounted on the sledge that are used to measure the horizontal field of view and are 75 cm apart.

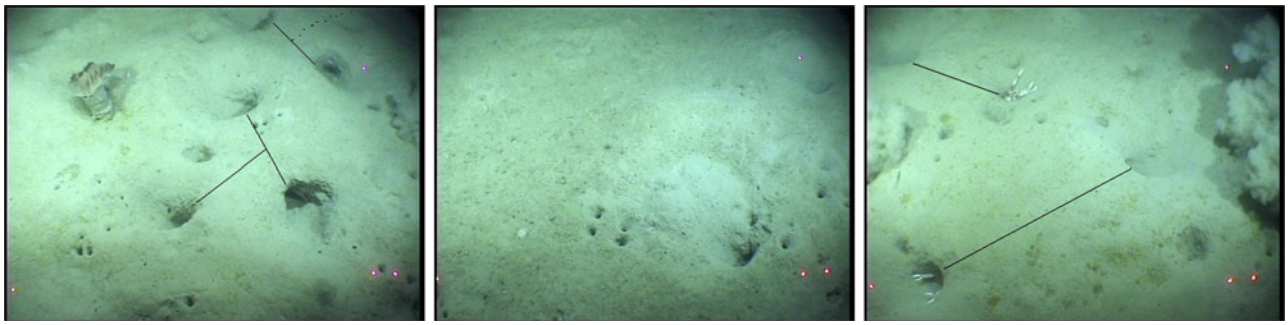

Fig. 4. Visual features experts use to classify nephrop burrows (adapted from [11]). The primary characteristics are crescent-shaped entrances arranged in a radial pattern around a common center (marked with the black lines in the left image). Experts also look for ejected sediment in the area around burrow openings (see the lighter area of sediment to the left of the large opening in the middle image) and the presence of nephrops in openings (two are partially visible in the right image).

is easier than in a video, mosaics allow the generation of a record of counted burrows for relatively little additional effort. In addition, mosaics can be used as a medium for further automation of the counting effort. In [7], we showed how burrow openings could be detected from mosaics of station videos. It would also be possible to use the burrow annotations to determine additional scientifically relevant information on the composition and size of burrows. The mosaicing algorithm we developed for this application is described in Section IV.

III. Brief Overview of Mosaicing

Mosaicing of sets of images has long been a topic of interest in the computer vision community [14]–[16]. Although the primary interest in mosaicing has been the creation of wide-angle panoramas from pictures taken by hand-held cameras [16], mosaics have been used in a variety of applications including seabed mapping [17]–[21]. At a basic level, mosaicing can be defined as the process of summarizing the content of a set of images of the same scene into a single image. Therefore, the key tasks in mosaicing are image registration, to establish the correspondences between images, and image rendering, to ensure that the transitions between images that form the mosaic are imperceptible.

A. Image Registration

Robustness and accuracy of image registration is critical for mosaicing algorithms. As image-to-image homographies have to be chained together to establish the positions of each frame in the mosaics, even small errors in the individual homographies can significantly impact the mosaic. The choice of model for the homographies describing these correspondences is a key choice in registration. These homographies range from simple two-parameter translation-only models to eight-parameter projective homographies capable of modeling perspective effects.[2] Furthermore, it is possible to use knowledge of the camera or the scene to reduce the number of degrees of freedom in a projective homography, for example, assuming the camera remains fixed and only rotates about its center [14], [16]. Although more elaborate models can describe a wider range of camera movement, reducing the number of degrees of freedom generally makes registration more robust.

Given the success of feature transforms such as SIFT [23] and SURF [24] at establishing point correspondences between images in challenging conditions, their use in image registration algorithms has become predominant. The homography parameters can be estimated by minimizing an energy function defined on the point correspondences using robust least squares approaches such as the Random Sampling Consenus (RANSAC) algorithm [16], [20], [25] or least median of squares [18], [26]. To further enforce global consistency of homography estimates, many mosaicing algorithms jointly optimize the homography parameters for all frames [14]–[16], [20], [21], [27], [28].

[2]A detailed description of different types of homographies relevant to mosaicing is given in [22, ch. 2, 13, and A.7].

B. Image Rendering

Once all of the images have been registered to a common coordinate system, a decision must be made on how to render parts of the mosaic where data are available from multiple images. Due to differences in illumination, movement of objects and misregistration of images, simply overwriting existing data in the mosaic as each frame is added will create visible seams along image boundaries. In [29], Prados *et al.* outline the key three stages in image fusion: minimizing the effects of illumination differences; finding the transition paths between the images; and blending of images across the transition paths.

1) Illumination correction: Illumination differences consist of both global and local illumination differences caused by different exposure times for each image and image vignetting or nonuniform lighting [16], [30].
2) Transition path selection: A simple solution to this problem is to choose the transition paths based on the distance of a point to the center of each frame. However, where the paths cross locations with registration misalignments or moving objects, this approach can lead to blurring or ghosting artefacts in the mosaic. To eliminate these differences, mosaicing algorithms sometimes employ an optimal seam finding technique (e.g., using graph cuts [29], [31]) to place the transition path in areas where there is minimal color differences between the images.
3) Image blending: Blending is used to avoid visible transition boundaries in the mosaic. Two broad approaches to blending have been employed in the literature. The first, referred to here as multiband blending, uses an invertible multiscale image decomposition such as Laplacian pyramids [32] or discrete wavelet transforms [33]. The idea is to adjust the width of the transition band at each scale so that high-frequency detail is not blurred by the blending. The second approach, known as gradient-domain blending, performs blending on image gradients and subsequently recovers image colors by solving a sparse system of linear equations [29], [31], [34].

C. Mosaicing of Underwater Video

Mosaicing has long been a popular topic of research in the underwater vision community for the mapping of the seafloor from cameras mounted on remotely operated vehicles or autonomous underwater vehicles (AUVs). However, in recent years, underwater mosaics have been used to monitor the condition of the seafloor and the health of benthic flora and fauna [9], [35]–[38].

There are a number of factors that necessitate a difference in approach when mosaicing underwater videos compared with generating panoramas from photographs taken by hand-held cameras. First, capturing video underwater poses many challenges for image quality. Attenuation, absorption, and scattering of light mean that underwater video has a limited range of view before features become blurred and have poor color contrast [29], [39]. As artificial lighting is necessary, nonuniform illumination across a video frame is also common [19], [20].

Unlike panorama generation, where it is usually assumed that cameras are only rotating, in the underwater environment, the camera motion is also translational. In general, this means that a full 3-D reconstruction is necessary to faithfully map the seabed [40]. However, under the assumption that the surface being mapped is planar, camera motion can also be modeled using a 2-D planar projective homography [22, ch. 13]. Although some of the earlier approaches to underwater mosaicing estimate full eight-parameter homographies [18], [20], [41], most state-of-the-art approaches try to use domain-specific knowledge to minimize the number of degrees of freedom to ensure better registration robustness. Assuming camera rotation and translation and that the seabed is a plane, a homography with 6 DOF can be employed [27], [28]. If camera motion is translational only and the direction of translation is parallel to the seabed, then the homography is an elation with 4 DOF [22, ch. 13 and A.7]. For cameras that are downward facing, a four-parameter similarity transform can be used [19], [21], [39].

In general, underwater mosaics are composed of many more images than a camera panorama. It is possible for an underwater mosaic to be composed from thousands or tens of thousands of images. This poses an increased premium on robustness due to the accumulation of errors in homographies. Furthermore, an increased computational cost is implied. Where navigation data are available, it can be incorporated into registration to improve robustness [28], [42] As an exhaustive search for pairs of frames for homography estimation is quadratic in the number of frames, sate-of-the-art techniques use domain-specific knowledge to make this process more efficient. As image data are generally captured by video cameras, homography estimation between successive frames can be used to estimate an initial camera trajectory than can be used to detect additional overlapping pairs of frames [20], [27], [28]. In [21], Elibol *et al.* outline a fast tentative initial trajectory estimation for unordered sets of images. More recently, researchers [43], [44] have also started to explore hierarchical mosaic generation from shorter submosaics as a means of improving robustness.

IV. Mosaicing Algorithm

A. Motivation

Fig. 5 shows frames from two different survey stations and demonstrate many of the visual aspects of the entire set. First of all, due to the artificial light source, a vignetting-like artefact is present. This is more pronounced at the top of the frame as these points are further away due to the forward and downward facing camera. Another consequence of the forward facing camera is the perspective effects induced by the forward camera motion. These sequences are captured in an underwater environment with a fair degree of turbidity. Hence, features at the top of the frame are noticeably more blurred in appearance than those at the bottom. Finally, the videos are stored in a compressed format and are often interlaced. Compression artefacts such as blocking artefacts and banding are also visible in the video frames.

Besides quality related issues, computation time is a crucial consideration. Given that station videos are typically 10-min long, a video at 30 frames/s will have 15 000 frames. Approximately ten stations are performed per day. Although real-time estimation of the mosaics is not a requirement, the algorithm

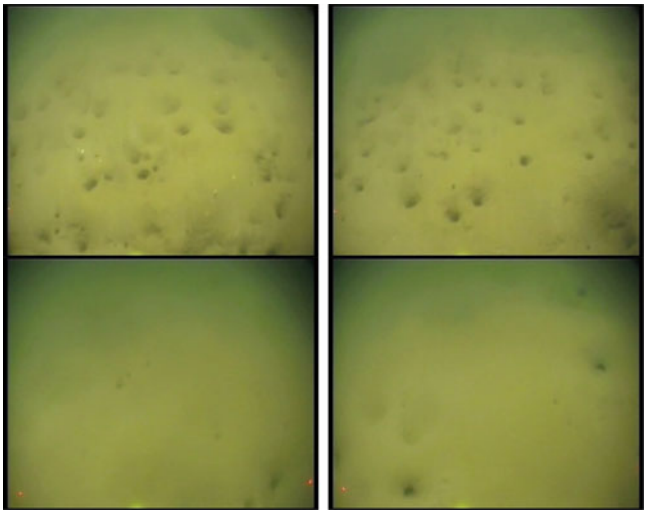

Fig. 5. Each row shows two frames from separate test sequences. The first sequence (top row) has a higher opening density and also better image sharpness than the second. Therefore, the second sequence is a more difficult sequence to register.

must be capable of generating the mosaics overnight to ensure that counting can take place on the boat.

For all of these reasons, the mosaics produced by general purpose mosaicing algorithms are insufficient for this application. Fig. 6 shows mosaics for the first test sequence from Fig. 5 generated by the Autostitch software [45] based on the work of Brown and Lowe [16] and the Microsoft Image Composite Editor (ICE) [46]. Even without examining the mosaics in fine detail, it can be seen that there are problems with both approaches. Examining the autostitch mosaic, it can be seen that there are a number of points where the perspective effects of the homographies are not estimated correctly, as one would expect even perspective over the sequence given the camera motion. Furthermore, the burrow openings are blurred in appearance. On the other hand, ICE shows a nonexistent camera rotation in the mosaic and exhibits an exaggerated color shift between the bottom and top of the mosaic. Furthermore, neither software is able to mosaic all the frames of the second test sequence correctly due to the lack of detail in the sequence.

Another issue with existing techniques is their large execution times. For example, both ICE and Autostitch take a long time[3] to mosaic the 900 frame sequences shown in Fig. 5. Therefore, it would not be feasible to mosaic 10 stations of 10 min duration in the required overnight time period. The main reasons for the high execution times of these methods are the global optimization of the registration parameters and the techniques for finding overlapping frames which has a complexity of $\mathcal{O}(n^2)$. Although underwater mosaicing algorithms [21], [28] propose optimizations of these processes, further improvements are required in this case given the exceptionally large number of frames and computational requirements.

Therefore, a new approach to mosaicing is needed. The key requirements are that the algorithm must be fast, robust, and

[3]Microsoft ICE takes about 1 h and Autostitch approximately 3 h to process the sequence on a PC with an 3.5-GHz Intel Xeon processor and 16 GB of RAM.

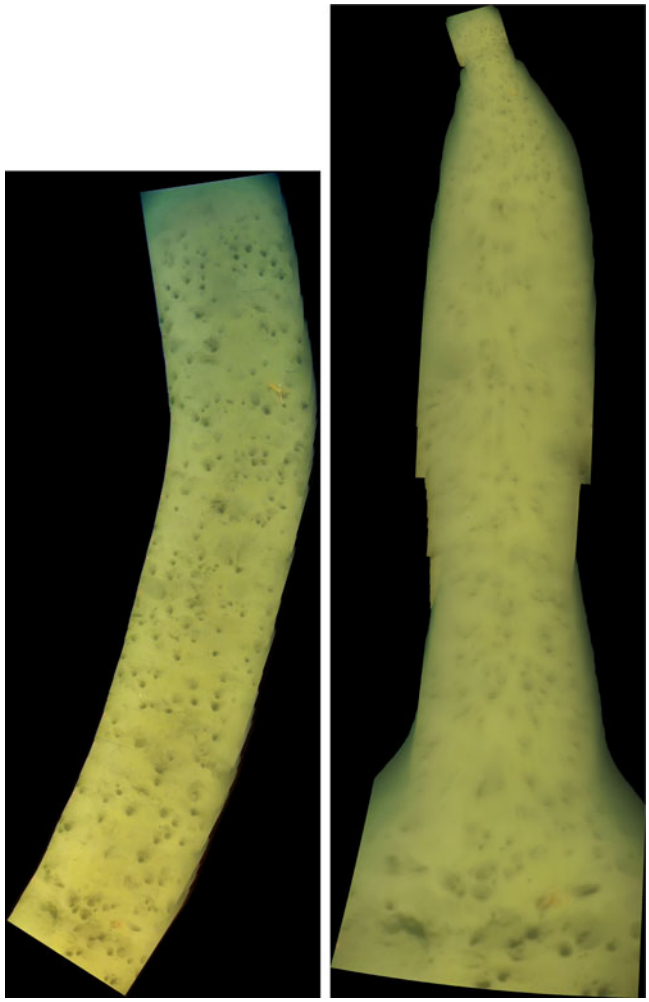

Fig. 6. Result of mosaicing the first test sequence shown in Fig. 5 with Microsoft ICE (left) and Autostitch (right).

must preserve all features relevant to the detection of burrows. To allow for robust estimation of motion parameters over a large number of frames, a translational motion model is used. Although it is not a sufficient model for a geometrically accurate rendering of the seabed, it preserves well the size and spatial relationships between openings from the video. Since the camera moves in a uniform direction, we find that homography estimation is only necessary between consecutive frames. Global optimization of frame-to-frame homographies is not performed to reduce computation. However, to improve robustness of the motion parameters to noise, we employ a maximum *a posteriori* (MAP) estimation framework which incorporates a temporal smoothness prior on the parameters. This prior acts like an adaptive autoregressive (AR) filter with increased smoothing when the number of feature matches is small. We adopt a multiband image blending scheme inspired by Burt and Adelson [32] to resolve overlap regions. However, we modify the weighting between frames to take account of the fact that most of the increased sharpness is at the bottom of frames. This implies that the mosaic is built from narrow horizontal strips from each frame. This also satisfies the counting protocols used by

marine biologists who count burrows that pass a horizontal line toward the bottom of the video frame [4], [13].

B. Registration

The problem of registration is that of estimating a homography between each frame and the mosaic. Given the position in frame i of a point k, $\mathbf{p}_{k,i}$, this mapping can be described using homogenous coordinates by

$$\mathbf{p}_{k,i} = H_{i,\mathrm{M}} \times \mathbf{p}_{k,\mathrm{M}} \tag{1}$$

where $H_{i,\mathrm{M}}$ is the 3×3 homography matrix mapping the position of point k on the mosaic, $\mathbf{p}_{k,\mathrm{M}}$, to its position in frame i.

A standard way solving this problem is to choose a reference frame from the video and map all other frames to the reference. Choosing frame 1 as the reference implies that $H_{1,\mathrm{M}}$ is the identity matrix. Instead of using bundle adjustment to jointly optimize the set of mosaic-to-frame homographies, we concatenate frame-to-frame homographies. Given $H_{b,a}$ is the homography mapping a point in frame a to a point in frame b

$$H_{i,\mathrm{M}} = H_{i,1} = \prod_{j=1}^{i-1} H_{i-j+1,i-j}. \tag{2}$$

As the accuracy of the estimate for $H_{i,\mathrm{M}}$ depends on the accuracy of the frame-to-frame homographies for the preceding video frames, robustness of the frame-to-frame homographies is critical. Therefore, we adopt a feature-based approach and perform a MAP estimation of the homographies from a set of inlying feature matches. This allows us to robustly estimate homographies in variable illumination conditions and gives a degree of robustness to frames where there are relatively few feature matches.

1) Feature Selection and Matching: Feature-based approaches attempt to establish a set of point correspondences, $\{(\mathbf{p}_{1,j}, \mathbf{p}_{1,i},), (\mathbf{p}_{2,j}, \mathbf{p}_{2,i}), \ldots\} \in \mathcal{P}_{j,i}$, between two frames i and j. We use both the SIFT [23] and SURF [24] algorithms to generate a set of features in each frame and the features are matched by comparing their respective descriptors using the technique described in [47]. Both algorithms are widely used in the state of the art for both general purpose and underwater mosaicing (e.g., [14], [16], [21], [28], and [44]) due to their excellent robustness to variations in illumination, scale, and orientation. Thus, two sets of key-point matches (one from each feature transform) are generated for each pair of frames.

It is important to consider whether the video is interlaced before forming feature extraction as most frame extraction tools will create images containing both fields in the even and odd rows of the frame images. This introduces jagged edge artefacts during camera motion which would lead to many erroneous keypoint detections. Therefore, for interlaced video, we only use the odd lines of the image for feature detection (i.e., an image that has full horizontal but only half vertical resolution). The y coordinates of the extracted keypoints are represented in the coordinate system of the original full-resolution image.

Another source of erroneous keypoint detections is the presence of banding artefacts. Banding artefacts manifest as a staircase intensity profile in areas with slowly changing gradients and are caused by quantization of the low-frequency bands of the discrete cosine transform during video compression. This results in a number of erroneous feature matches in the top-left and top-right corners of the frames where the banding artefacts are most noticeable. Generally, the positions of these keypoints are quite stable from frame to frame. This can cause a failure of robust homography estimation when there are relatively few genuine feature matches between frames. There are a number of approaches in the literature for suppressing banding artefacts through debanding [48] or dithering [49]. However, since turbidity of the water causes seabed features to be often severely blurred at the top of the frame (see Fig. 5), feature matching is instead suppressed in a horizontal band spanning the top portion of each image.

2) Homography Estimation: The homography between consecutive frames i and $i+1$ is now estimated using the set of keypoint matches between the frames. Given the set of keypoint matches $\mathcal{P}_{i+1,i}$ and the previous homography estimate $H_{i,i-1}$, the task is to find the homography $\hat{H}_{i+1,i}$ that maximizes $p(H_{i+1,i}|\mathcal{P}_{i+1,i}, H_{i,i-1})$. Using Bayes' theorem, and dropping the subscripts for brevity, the posterior can be factorized into a data likelihood $p_l(.)$ and a temporal smoothness prior $p_s(.)$ as follows:

$$p(H|\mathcal{P}, H_0) \propto p_l(\mathcal{P}|H, H_0) \times p_s(H|H_0) \tag{3}$$

where $H_0 = H_{i,i-1}$.

For a full-planar projective homography, H is a 3×3 matrix with 8 DOF. As explained in Section IV-A, our algorithm uses a translational-motion-only model, and for usability purposes it is restricted further to vertical translation only. Therefore, $H_{j,i}$ is of the form

$$H_{i+1,i} = \left(\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & t_{\mathrm{y}} \\ 0 & 0 & 1 \end{bmatrix} \right) \tag{4}$$

where t_{y} is the vertical translation parameter.

a) Likelihood: Under the assumption that the set of keypoint matches $\mathcal{P}$ is independent of H_0, a general form of the likelihood is given by a Gaussian distribution as

$$p_l(\mathcal{P}|H, H_0) \propto \exp\left(-\frac{f(\mathcal{P}, H)}{2\sigma_l^2}\right) \tag{5}$$

where $f(\mathcal{P}, H)$ is a nonnegative cost function and σ_l^2 is some variance. A good overview of the various cost functions for a planar projective homography is given in [22, ch. 4]. Given a vertical translation motion model, we choose a cost function based on the geometric distance between the inhomogeneous coordinate of the point $\mathbf{p}_{k,i+1}$ and its predicted position given the homography $H\mathbf{p}_{k,i}$. Using $d(\mathbf{p}, \mathbf{q})$ to represent the distance between the inhomogeneous form of generic points $\mathbf{p}$ and $\mathbf{q}$, the cost function is given by

$$f(\mathcal{P}, H) = \sum_k d(\mathbf{p}_{k,i+1}, H\mathbf{p}_{k,i})^2. \tag{6}$$

b) Prior: The aim of temporal smoothness prior is to force H to be similar to H_0. However, since both H and H_0 are redundant up to a scale, their values must be normalized before the similarity between them can be calculated. We normalize both matrices by scaling them such that their bottom right elements are equal to 1. Rearranging the normalized matrices as column vectors $\tilde{\mathbf{h}}$ and $\tilde{\mathbf{h}}_0$, the prior distribution is expressed using a Gaussian distribution as

$$p_s(H|H_0) \propto \exp\left(-\frac{\left\|\tilde{\mathbf{h}}-\tilde{\mathbf{h}}_0\right\|^2}{2\sigma_h^2}\right). \tag{7}$$

c) Optimization: Taking negative log likelihoods of both sides of (3) and incorporating the variances σ_l^2 and σ_h^2 into a single nonnegative smoothness parameter Λ_s ($\Lambda_s = \sigma_l^2/\sigma_h^2$), the MAP problem is transformed into an energy minimization problem where the energy is

$$E(H) = f(\mathcal{P}, H) + \Lambda_s \times \left\|\tilde{\mathbf{h}}-\tilde{\mathbf{h}}_0\right\|^2. \tag{8}$$

For the expression of $f(\mathcal{P}, H)$ in (6), $E(H)$ is a quadratic equation H and hence can be minimized using standard linear least squares techniques. For a vertical translation model, the least squares estimate of t_{y} in (4) is given by

$$\hat{t}_{\mathrm{y}} = \frac{\sum_k (y_{k,i+1} - y_{k,i}) + \Lambda_s \times t_{\mathrm{y},0}}{\sum_k 1 + \Lambda_s} \tag{9}$$

where y_* is the vertical coordinate of $\mathbf{p}_*$ and $t_{\mathrm{y},0}$ is the vertical translation parameter for the previous pair of frames. The RANSAC [25] algorithm is used to ensure that outliers are removed from the feature matches before the optimization described in (9) is carried out.

A property of this solution for t_{y} is that the amount of smoothness provided by the prior is dependent on the number of feature matches between the pair of frames. More weight will be placed on the smoothness prior when the number of feature matches are low. This is a desirable property as it matches our intuition that we have more confidence in the homography estimate gained solely from the feature matches when the number of matches is greater. The solution in (9) can be thought of as an adaptive AR filter on the time series of homography parameters calculated solely on the feature matches. This property also holds for motion models with more degrees of freedom as long as $f(\mathcal{P}, H)$ is not normalized by the number of feature matches between frames.

C. Rendering

Our blending framework is based on the multiband blending framework proposed in [32]. It consists of selecting the optimum regions from each frame and multiband blending itself. A key difference of our method is how the blending weights of each frame are selected. Of the techniques employing multiband blending, typically a center-weighting technique is applied [14, ch. 9]. This assumes that the best detail is located at the center of the frame. Center-weighting is an effective means of dealing with vignetting as only the well-lit central regions are kept in the mosaic [16]. Many underwater mosaicing algorithms [20], [27], [42] also use forms center-weighting as underwater mosaics are typically built from downward facing cameras and so the illumination pattern caused by artificial lighting resembles a vignetting artefact. However, cameras in UWTV surveys are typically slightly forward-facing and so the illumination pattern in video frames is not symmetric and turbidity of the water can cause blurring of texture in the top portion of the frame. Furthermore, counting protocols [4], [13] require burrows to be counted as they pass through a horizontal line toward the bottom of the frame.[4]

Therefore, it is desirable to use the bottom portion of each frame in the mosaic. To meet these requirements, we introduce a center-weighting scheme that uses an arbitrary center location which is chosen toward the bottom of the frame. Like the center-weighting scheme in [16], the center pixel is given a weight of 1 which drops to 0 at the boundaries of the frame.

In our framework, the mosaic is built up one frame at a time with the current frame being blended with the existing mosaic. The first stage is the computation of the color values of the registered frame according to the computed homographies. This can be achieved using an interpolation scheme (e.g., bilinear or bicubic interpolation) to retrieve color values at noninteger pixel locations. Again, special consideration is required for interlaced frames and interpolation is performed on the same interlaced field (odd or even) used for feature selection.

Given the intensity of a color channel of the mosaic $M(\mathbf{x})$ and a registered frame $F(\mathbf{x})$, the updated mosaic $M'(\mathbf{x})$ is obtained by multiband blending of Laplacian decompositions of $M(\mathbf{x})$ and $F(\mathbf{x})$. The Laplacian decomposition of $F(\mathbf{x})$ consists of N bandpass subbands $F_{\mathrm{L},0}(\mathbf{x})$, $F_{\mathrm{L},1}(\mathbf{x})$, …, $F_{\mathrm{L},N-1}(\mathbf{x})$ and a lowpass subband $F_{\mathrm{L},N}(\mathbf{x})$. The Laplacian decomposition is calculated by first computing a similar sized Gaussian decomposition. The Laplacian subband at level n is then given by

$$F_{\mathrm{L},n}(\mathbf{x}) = \begin{cases} F_{\mathrm{G},n+1}(\mathbf{x}) - F_{\mathrm{G},n}(\mathbf{x}), & n = 0, \ldots, N-1 \\ F_{\mathrm{G},N}(\mathbf{x}), & n = N \end{cases} \tag{10}$$

where the nth-level subband of the Gaussian decomposition is

$$F_{\mathrm{G},n}(\mathbf{x}) = F_{\mathrm{G},n-1}(\mathbf{x}) * g(\mathbf{x}) \tag{11}$$

with $F_{\mathrm{G},0}(\mathbf{x}) = F(\mathbf{x})$ and where $g(\mathbf{x})$ is a zero-phase lowpass FIR filter. A similar Laplacian decomposition is constructed for $M(\mathbf{x})$.

The updated color channel of the mosaic after multiband blending is defined by

$$M'(\mathbf{x}) = \sum_{n=0}^{N} \left(w_n(\mathbf{x})\, F_{\mathrm{L},n}(\mathbf{x}) + (1 - w_n(\mathbf{x}))\, M_{\mathrm{L},n}(\mathbf{x})\right). \tag{12}$$

This requires the definition of blending weights $w_n(\mathbf{x})$ for each level of the decomposition. The multiband blending weights for

[4]Burrows that pass through the left or right boundaries of the frame are not counted.

each frame are initialized from the center-weighting maps for the frame $c_{\mathrm{F}}(\mathbf{x})$ and the mosaic $c_{\mathrm{M}}(\mathbf{x})$ as follows:

$$w(\mathbf{x}) = \begin{cases} 1, & c_{\mathrm{F}}(\mathbf{x}) > c_{\mathrm{M}}(\mathbf{x}) \\ 0, & \text{otherwise.} \end{cases} \tag{13}$$

As noted in [32], the strength of multiband blending is the use of different transition widths in the blending weights at each level. Transition widths for high-frequency image detail are low while those for low-frequency content are larger. This ensures that seams caused by a step change in intensity across a transition boundary are removed with minimal perceived blurring. Therefore, following [16], we define the blending weights for level 0 as

$$w_0(\mathbf{x}) = w(\mathbf{x}) * g(\mathbf{x}) \tag{14}$$

and the blending weights for level n as

$$w_n(\mathbf{x}) = w_{n-1}(\mathbf{x}) * g(\mathbf{x}). \tag{15}$$

Finally, the center-weights for the mosaic are updated according to

$$c'_{\mathrm{M}}(\mathbf{x}) = \begin{cases} c_{\mathrm{F}}(\mathbf{x}), & c_{\mathrm{F}}(\mathbf{x}) > c_{\mathrm{M}}(\mathbf{x}) \\ c_{\mathrm{M}}(\mathbf{x}), & \text{otherwise.} \end{cases} \tag{16}$$

V. Results

A. Initial Testing

The mosaicing algorithm was initially trained and tested on 20 test sequences taken from UWTV surveys off the west coast of Ireland and in the Irish Sea. The videos vary in length from 30 s to 3 min. The videos are captured by a PAL SD camera and are encoded as H.264 video at a bit rate of approximately 500 kb/s. The videos are all interlaced and vary in frame rate from 10 up to 29.97 frames/s. All the video frames contain a small black border and so the frames are cropped by 20 pixels on all 4 sides before they are mosaiced.

1) Implementation Details: The prototype of the algorithm to conduct the experiments outlined in this paper is implemented in MATLAB [50]. The inbuilt functions *detectSURFFeatures* and *extractFeatures* are used to generate SURF keypoints and their descriptors, respectively. The implementation of SIFT algorithm developed by Vedaldi [51] is used to detect the SIFT keypoints and descriptors. As turbidity of the water means that image detail is often blurred and poorly contrasted, both feature detectors are biased to encourage more keypoint detections. The value of the *MetricTheshold* parameter is set to 5 and the region of interest (ROI) is set to exclude a horizontal band spanning the top 100 rows of each image.[5] For the SIFT features, the *Threshold* parameter is reduced from its default to 0.002 and we set the *NumOctaves* parameter to 4 and the *NumLevels* parameter to 3. Again features are not detected in the top 100 rows of the image. All other parameters in both detectors are set to their default values. SURF features from pairs of frames are matched

[5]This corresponds to the 18.7% of the vertical resolution of the frames in our test set. If the frame is interlaced, only the top 50 rows of the odd or even fields are excluded.

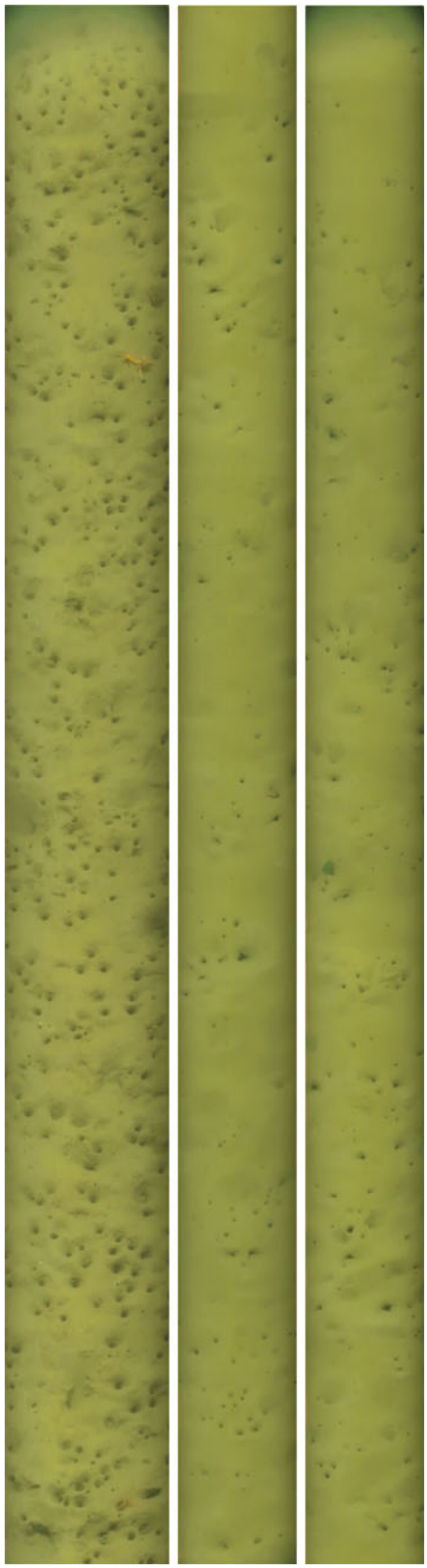

Fig. 7. Mosaics for the two test videos shown in Fig. 5. The left image shows the mosaic for the first test sequence which is 36 s long. For visualization purposes, the mosaic for the second test sequence (1 min long) has been split into two halves in the middle (bottom half) and right (top half) images.

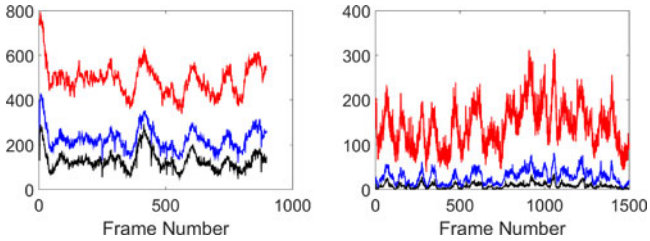

Fig. 8. Plots of number of feature detections (red), feature matches (blue) and inliers after RANSAC (black) against frame number for test sequences 1 (left) and 2 (right). Overall, there are few fewer feature detections, matches, and inliers for the second test sequence than the first test sequence.

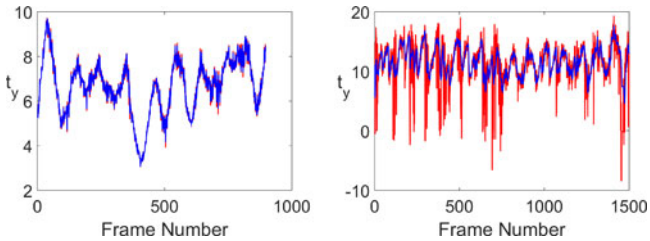

Fig. 9. Plots of translation parameter t_y against frame number for test sequences 1 (left) and 2 (right). The blue plots represent the result with the default value of $\Lambda_s = 15$ and the red plot represents the case where the smoothness prior is turned OFF ($\Lambda_s = 0$). The effective smoothing in sequence 2 is much greater than for sequence 1.

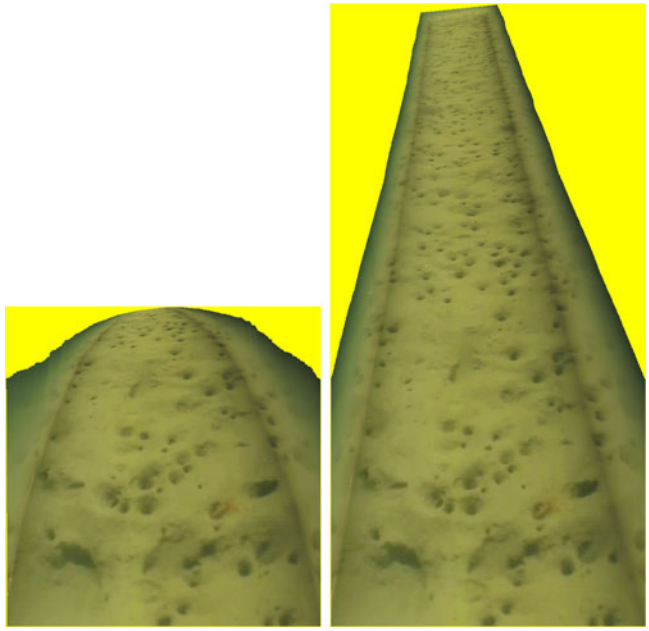

Fig. 10. Mosaics generated for test sequence 1 using a planar projective homography with 8 DOF (left) and an elation with 4 DOF (right). The perspective present in the mosaic could be compensated for by using the appropriate mosaic coordinate system.

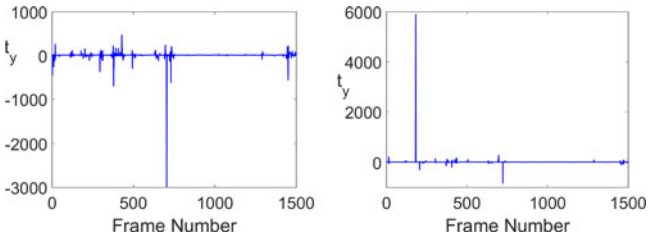

Fig. 11. Plots of the vertical translation parameter against frame number for test sequence 2 for both a full-planar projective (left) and elation (right) motion models. In these plots the temporal smoothness prior has been turned off. For each frame the homography matrix has been normalized so that the bottom right element is equal to 1. At this scale, well behaved homography parameters are close to 0 (as in Fig. 9) and values visibly deviating from zero constitute significant noise in the parameters. Note the high level of noise corresponding to frames with low numbers of features matches (see Fig. 8).

using the *MatchFeatures* function of MATLAB and SIFT features use the *siftmatch* function provided in [51]. Both feature matching functions use their default values.

In the robust homography estimation algorithm, 200 trials of RANSAC are performed. In each trial, three feature matches are randomly sampled and a least squares estimate of the vertical translation motion model is found. If there are fewer than three feature matches, then the homography parameters are set to those of the previous frame and the homography estimation stops. The inliers according to this translation are determined by finding the subset of matches that are located within 2 pixels of their predicted positions. The chosen set of inliers is the largest subset of inliers across all trials or the first trial where the number of inliers exceeds 75% of the total number of feature matches. From this set of inliers, a final homography is estimated according to (9).

For the multiband blending, the center chosen for the center-weighting is the midpoint of the row 50 rows up from the bottom of the frame.[6] A Gaussian filter is chosen for $g(\mathbf{x})$ that has a size of 30×30 taps and has a σ value of $30/6.4$. The coefficients of the filter are normalized so that they sum up to 1. A value of 2 is chosen for N in the Laplacian decompositions.

2) Experimental Results: The mosaics for two of these test sequences (shown in Fig. 5) are shown in Fig. 7. The first test sequence is 30 s long at 29.97 frames/s while the second is 2.5 min long at 10 frames/s. The first sequence contains many burrow openings while the second is much more sparsely populated. Both sequences show moderate turbidity with the top of each frame noticeably more blurred than the bottom. As can be seen from Fig. 7, both sequences have been mosaiced successfully as all countable burrow openings in the sequences are preserved in the mosaic. Furthermore, the multiband blending algorithm has ensured that there are no visible seams in the mosaics.

A key property of the algorithm is the robustness of the registration algorithm when the number of feature matches is low and this strength is derived from the temporal smoothness prior. This is shown in Figs. 8 and 9. Fig. 8 shows the number of feature detections, matches, and inliers per frame for both of the highlighted test sequences. The lack of burrow openings and image detail generally in test sequence 2 results in significantly fewer inlying feature matches per frame (an average of 8.3 for sequence 2 compared to 129.2 for sequence 1). For a standard maximum likelihood homography estimation algorithm, this would lead to noisy estimates for the translation parameter (see Fig. 9). However, the action of temporal smoothness prior

[6]This corresponds to 9.4% of the vertical image resolution.

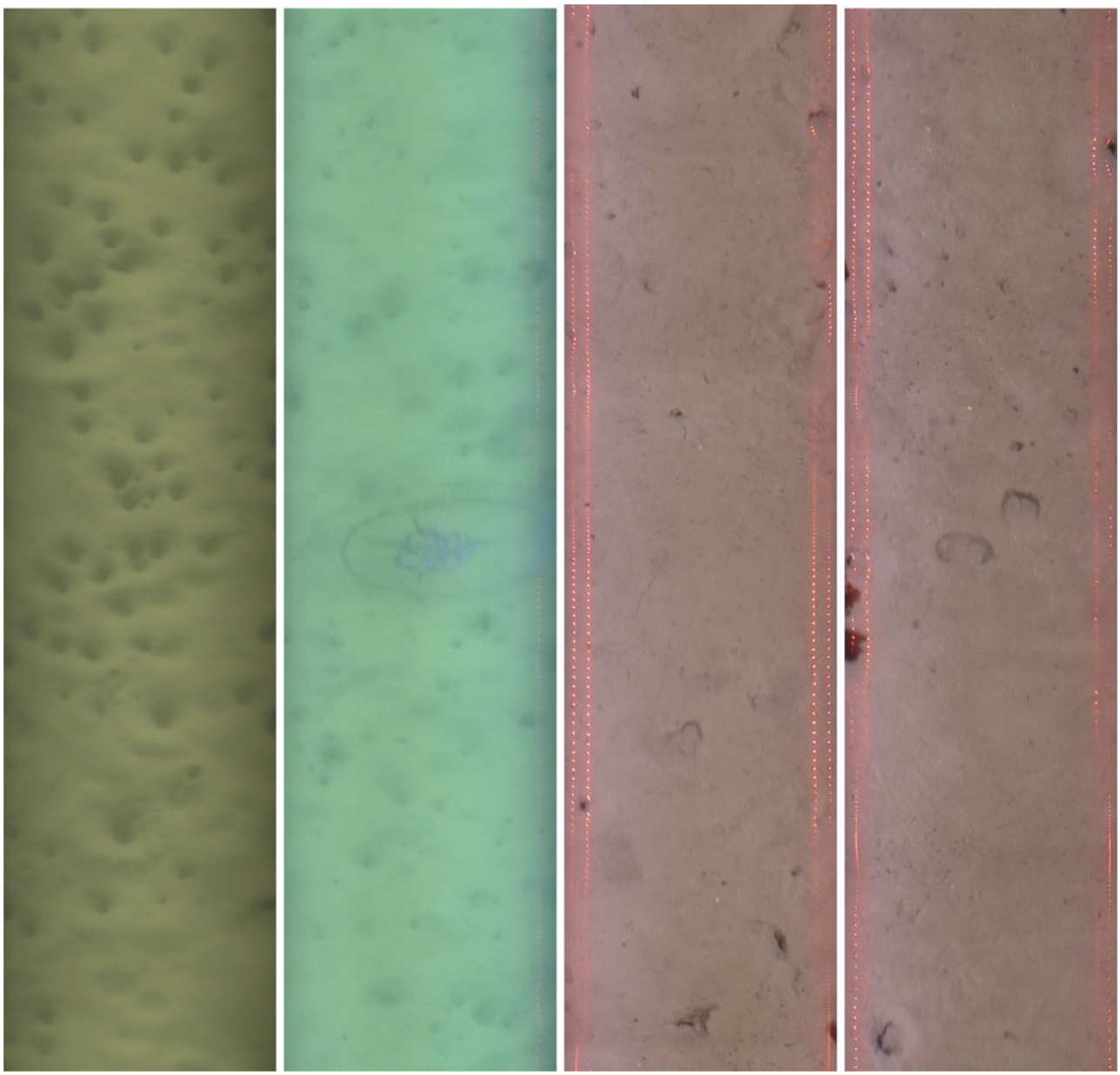

Fig. 12. Partial mosaics for the station videos depicted in Fig. 3. The trail of red dots in the mosaics on the right are caused by the lasers mounted on the sledges use to calibrate horizontal scale.

significantly reduces this noise as its strength is greater given the smaller numbers of inliers. In comparison, the prior has little effect on test sequence 1 as the number of inliers is much greater.

Computation time is much reduced when compared with off-the-shelf options such as Microsoft ICE or Autostitch. A MATLAB prototype of the algorithm takes approximately 12 min to process the 900-frame first test sequence, whereas Microsoft ICE took around an 1 h and Autostitch 3 h on the same PC. With a more efficient multithreaded implementation of our algorithm, the computation times of our algorithm could be reduced sufficiently for use on Nephrops Surveys.

A drawback of using the vertical translation motion model is that vertical scale is not the same as horizontal scale due to the perspective induced by the camera orientation. This distortion could be corrected by using a motion model that accounts for perspective. To assess the feasibility of using such transforms, we computed homographies for the test sequences using a full-planar projective homography as well as an 4 DOF elation, which is valid for a camera undergoing translation parallel to a planar surface. The Sampson error is used for $f(\mathcal{P}, H)$ in (8) and nonlinear least squares optimization is used to estimate the motion parameters. When a large number of feature matches are present, plausible homographies are estimated. This can be seen from the mosaics for test sequence 1 shown in Fig. 10. However, the difference in apparent perspective between both models indicates the presence of significant noise in the parameters of the full-planar projective homographies. The noise increases dramatically when the number of feature matches is low. This is evident in the values of the vertical translation

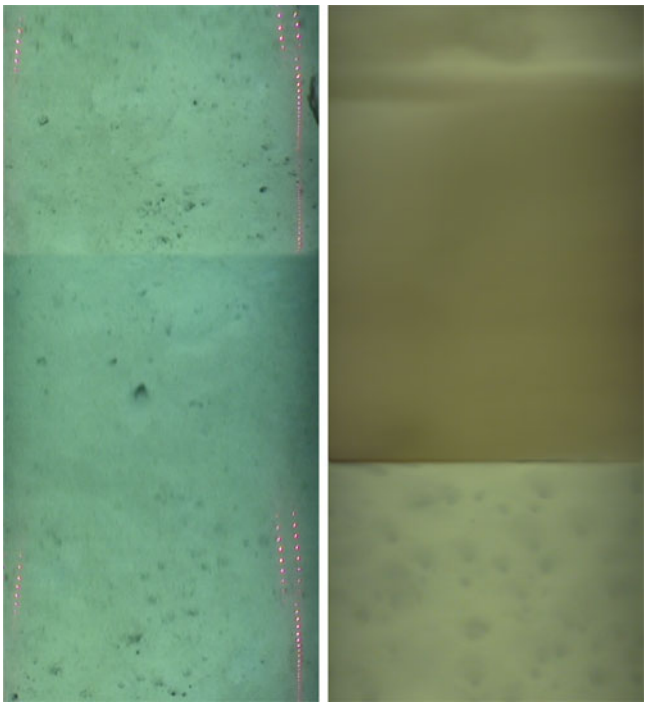

Fig. 13. (Left) A portion of a mosaic with an artefact introduced by wave motion periodically pulling the camera sledge off the seabed. (Right) A portion of a mosaic where the disturbed sediment has fully occluded the field of view of the camera.

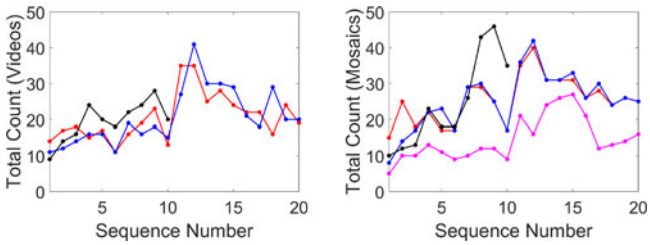

Fig. 14. (Left) A graph showing the count totals for the three counters using the existing video-based techniques. (Right) A graph depicting the total number of the annotated burrows in each mosaic. The counter represented by the black plots only returned counts for sequences 1 to 10. The purple plot on the right graph represents the total number of burrows that were common to each of the counters for that sequence.

parameters of the homographies estimated for test sequence 2 shown in Fig. 11. The temporal smoothness prior cannot suppress large levels of noise in these parameters without significant oversmoothing of the homography parameters. Due to the poor performance of projective homographies with low numbers of feature matches, we consider them unsuitable for this application.

B. Processing of 2014 Underwater Television Survey

The algorithm was tested on the video data set generated by the 2014 UWTV annual survey of the major Nephrop fishing grounds in Irish coastal waters. These surveys were conducted over 30 days on board the RV Celtic Voyager in FUs 15, 16, 17, 19, 20/21, and 22 (see Fig. 1). The video data were gathered from an analog PAL camera mounted on a sledge. The video is digitally captured in real time on the research vessel onto DVDs and is compressed using the MPEG2 codec at a bit rate of approximately 10 Mb/s. The spatial resolution of the captured videos is 720×576 and has a frame rate of 25 frames/s. All of the videos were interlaced.

Mosaics were generated for 365 different station videos. Most station videos are 10-min long with approximately 15 000 frames. As the DVD format meant that station videos did not align with the encoded video files, the start frame and end frame of each station video was recorded manually. A sample of mosaics generated from this data set are shown in Fig. 12. More videos and corresponding mosaicing results are shown online at http://www.mee.tcd.ie/~corrigad/?page_id=240. Results over the entire data set show that the algorithm performs well over terrain with varying densities of burrow openings and levels of turbidity. The same settings described in Section V-A1 are used for all sequences in the survey data.

A number of challenges arise on this data set that did not exist in the original test set. The most significant of these is the effect of wave action on the camera motion. As the camera sledge is attached directly to the research vessel, wave motion can cause the camera position to move up and down and to tilt. This camera motion induces changes in scale and perspective that cannot be modeled by the motion model used in the algorithm. However, the simplicity of the motion model ensures that wave motion does not significantly distort the visual structure of the mosaics and the impact is limited to localized distortions in vertical scale and discontinuities in illumination and sharpness (see Fig. 13). Most importantly, counting is not performed when such motion occurs. Another challenge is the presence of disturbed sediment, often caused by contact of the sledge with the seabed, which can cause the partial or total occlusion of the seabed in the video frames. However, the simplicity of the motion model along with the use of RANSAC and the smoothness prior mean that reasonable homography parameters can still be obtained (see Fig. 13).

VI. Discussion

The generation of mosaics has a large potential benefit to the efficiency and scientific value of UWTV surveys for stock assessment benthic species including Nephrops. In the short term, they have the potential to change the manual counting procedure currently employed. Using mosaics allows burrows to be annotated and, thus, can create a record of counted objects that is not currently possible with existing techniques. This, in turn, can be used to improve the quality and consistency of counts. Furthermore, the annotations could be used to quantify other useful information that could be used to monitor population dynamics.

To gain an insight as to the potential impact of counting from mosaics instead of the videos, we asked three expert counters from three different organizations to independently count the number of burrows in the 20 sequences in the initial test set discussed above. Counts were initially performed on the videos

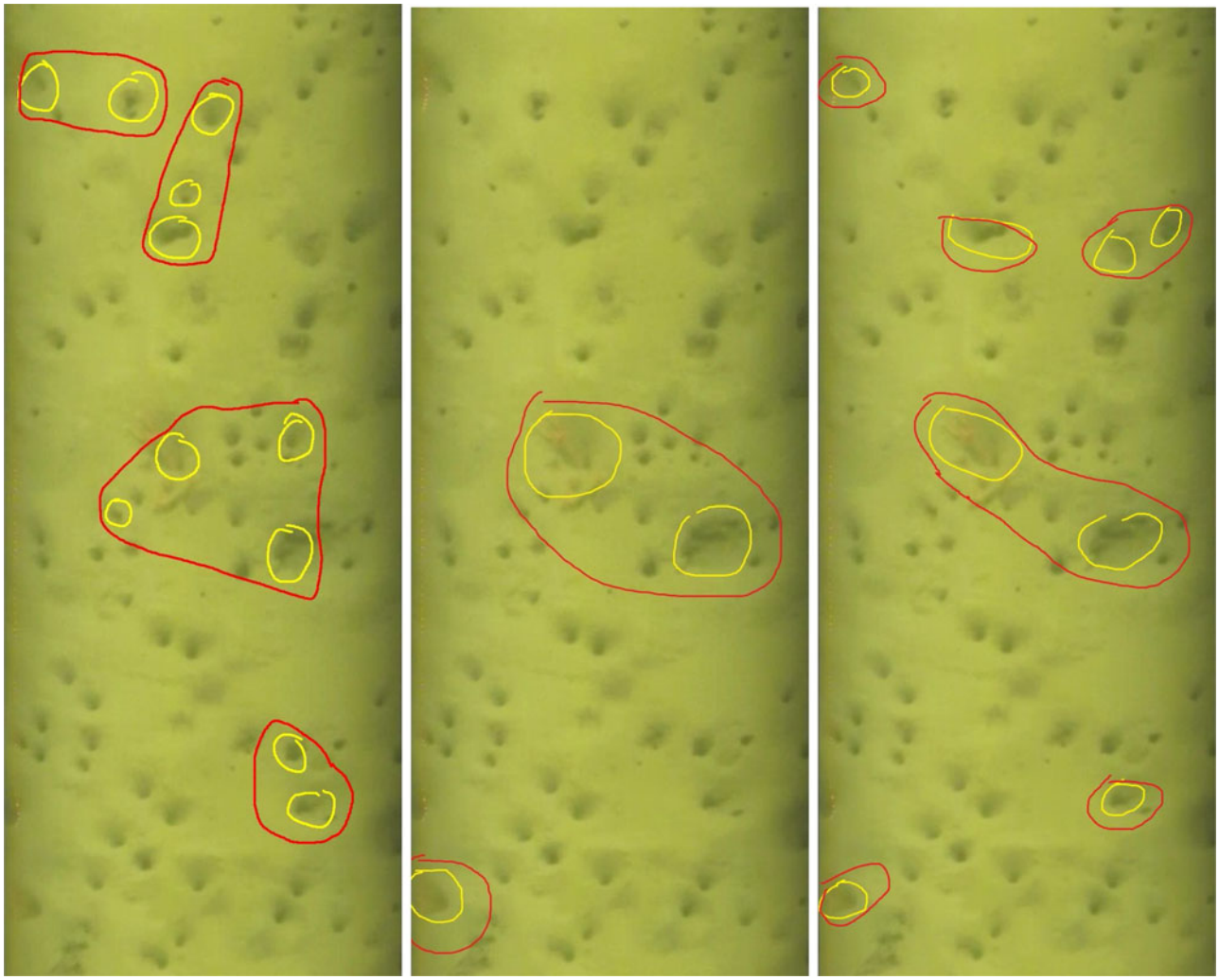

Fig. 15. Different annotations for the three counters for a common portion of one of the test sequences. Individual Nephrops burrow openings are circled in yellow while the burrows themselves are highlighted in red.

using the existing technique and then, after a short break, from the mosaics using Microsoft Paint to add the annotations. Two of the counters returned totals from all 20 sequences, while the third returned counts from the first 10 only. The results from the count are summarized in Fig. 14.

From the graphs in Fig. 14, it is immediately apparent that the count totals are higher for the mosaics than the videos (20% higher on average). It is possible that this is largely due to counters spending more time counting on the mosaics than the videos but it is also likely that burrows are being counted that would not have been counted using the existing technique. Although count totals are largely consistent between counters using both the mosaics and videos, when examined on a burrow-by-burrow basis our analysis shows that there is a significant discrepancy between counters. Of the approximately 26 burrows counted on average from the mosaics, only 14.5 of them are common to all of the counters. A visual example of this is shown in Fig. 15 for the labeling of the three counters for one of the test sequences. This shows the subjective nature of the existing counting protocols in a way not possible before. It also demonstrates the potential role of the mosaics in improving consistency amongst counters.

Of course the ideal scenario would be to have an automated population count based on machine learning. This applies to Nephrops but also other burrowing species and shellfish such as scallops. To date, automated counting of Nephrop burrows has achieved most interest [6]–[8] due to the commercial value of the Nephrops fishery. By summarizing the content of a video in a single image, the video object detection problem is transformed into a more tractable image object detection problem since there is no need to employ explicit tracking of objects of interest from frame to frame [6]. In [7], we showed how mosaics could be used in conjunction with simple supervised learning schemes to detect and count burrow openings.

Automatic counting of nephrop burrows remains an unsolved problem. Due to the complex appearance and relationships between openings of nephrop burrows, it is difficult to construct handcrafted feature sets to distinguish nephrop burrows from those of other burrowing species. Deep learning approaches have the ability to model complex visual characteristics of objects without the need for handcrafted features and clearly could

be used to count both nephrop burrow openings and entire burrows themselves. The biggest impediment to automated counting is the lack of a suitable ground truth for training. However, if counting from mosaics becomes the standard technique then large data sets will be generated for training as the manual annotations from counters could be used to create ground truth.

A. Role of Enhancement

As our goal was to design a robust mosaicing algorithm in difficult underwater environments, significant enhancement was not considered as part of the mosaicing pipeline. Applying enhancement as a preprocess to our algorithm could be used to improve the visual quality of the mosaics and possibly improve the performance of the mosaicing through the elimination of spurious feature matches. Enhancement could also improve the performance of subsequent automated burrow counting. In underwater video surveys, the main impacts on quality are related to the challenging imaging environment and also the compression applied to the videos. Perhaps water turbidity has the most significant impact on the appearance of the video particularly toward the top of the frame but can also vary depending on the local conditions at the survey site. There are a number of contrast enhancement techniques for underwater images to improve the loss of contrast caused by turbidity (see [52]–[54]) that could be used. Illumination correction (e.g., [55]) may also be of benefit but proper illumination can avoid serious uneven illumination. From our analysis, artefacts in video compression have an impact on the performance of mosaicing by causing spurious feature matches. These artefacts are related to blocking and banding introduced by quantization in the video codec and could be by applying deblocking or debanding filters (see [48] and [56]) to the frames before mosaicing. Therefore, enhancement should be considered as an important part of the overall mosaicing and burrow counting pipeline.

VII. Final Remarks

This paper presents an algorithm for mosaicing videos of the seabed used for UWTV surveys. It uses a vertical translational motion model to robustly register each frame to a common coordinate system using a simple concatenation of frame-to-frame homographies. A MAP framework is used for homography estimation which uses a temporal smoothness prior on the motion parameters that acts as an adaptive autoregressive lowpass filter on the homography parameters. This prior reduces noise in the parameter estimates, in particular, where the numbers of feature matches between frames are low. Rendering is performed using multiband blending scheme using a modified center-weighting that takes account of the burrow counting protocols and optimizes the visibility of the mosaic. The algorithm was tested on a data set from an annual UWTV survey and shows that the algorithm is sufficient for use in stock assessment of Nephrops. Developing an automated counting algorithm based on deep learning techniques is the research direction we are pursuing. The first stage of this will be to develop a counting solution based on the mosaicing algorithm that can be used on-board during the surveys. This will allow more verifiable and detailed manual counts as well as the creation of ground truth data sets for learning.

Acknowledgment

The authors would like to thank the officers and crew of *RV Celtic Voyager* and all scientists involved in the collection and review of the video footage. They would also like to thank C. Fitzpatrick for her work preparing the data and scripts for the mosaicing of the videos in the 2014 UWTV survey.

References

[1] S. Jennings, M. J. Kaiser, and J. D. Reynolds, *Marine Fisheries Ecology*. Hoboken, NJ, USA: Blackwell Science, 2009.

[2] Eurostat, "Landings of fishery products." [Online]. Available: http://ec.europa.eu/eurostat, Accessed on: Mar. 29, 2018.

[3] N. Campbell, H. Dobby, and N. Bailey, "Investigating and mitigating uncertainties in the assessment of Scottish Nephrops Norvegicus populations using simulated underwater television data," *ICES J. Marine Sci., J. du Conseil*, vol. 66, no. 4, pp. 646–655, 2009.

[4] "Workshop on the use of UWTV surveys for determining abundance in Nephrops stocks throughout European waters," Int. Council Exploration Sea, Rep. CM 2007/ACFM: 14, 2007.

[5] Eurostat, "Landings of fishery products in Ireland." [Online]. Available: http://ec.europa.eu/eurostat, Accessed on: Mar. 29, 2018.

[6] P. Lau, P. Correia, P. Fonseca, and A. Campos, "Estimating Norway lobster abundance from deep-water videos: An automatic approach," *IET Image Process.*, vol. 6, no. 1, pp. 22–30, 2012.

[7] K. Sooknanan, J. Doyle, J. Wilson, N. Harte, A. Kokaram, and D. Corrigan, "Mosaics for burrow detection in underwater surveillance video," in *Proc. OCEANS Conf., San Diego*, CA, USA, 2013, pp. 1–6.

[8] K. Sooknanan, J. Doyle, C. Lordan, J. Wilson, A. Kokaram, and D. Corrigan, "Mosaics for Nephrops detection in underwater survey videos," in *Proc. OCEANS Conf., St. John's*, NL, Canada, 2014, pp. 1–6.

[9] K. Sooknanan, A. Kokaram, D. Corrigan, G. Baugh, N. Harte, and J. Wilson, "Indexing and selection of well-lit details in underwater video mosaics using vignetting estimation," in *Proc. OCEANS Conf.*, Yeosu, South Korea, May 2012, pp. 1–7.

[10] Working Group on Nephrops Surveys, "Second interim report of the working group on nephrops surveys (WGNEPS)," Int. Council Exploration Sea, Tech. Rep. ICES CM 2014/SSGESST:20, 2014.

[11] The Marine Institute, "A guide to nephrops burrow identification and how to count nephrops burrow systems." [Online]. Available: http://www.marine.ie/Home/sites/default/files/MIFiles/Docs/FisheriesEco%systems/Nephrops%20norvegicus%20Burrow%20Identification.pdf

[12] C. Lordan *et al.*, "Report of the UWTV survey on the Aran, Galway bay and Slyne head nephrops grounds 2006," in *Proc. Workshop Use UWTV Surveys for Determining Abundance Nephrops Stocks Throughout Eur. Waters*, vol. CM 2007/ACFM:14, 2007, pp. 118–144.

[13] A. Weetman and N. Campbell, "Protocol for establishing a nephrops burrow abundance using under water video," in *Proc. Workshop Use UWTV Surveys Determining Abundance Nephrops Stocks Throughout Eur. Waters*, 2007, vol. CM 2007/ACFM:14. 2007, pp. 170–182.

[14] R. Szeliski, *Computer Vision: Algorithms and Applications*. New York, NY, USA: Springer-Verlag, 2010.

[15] H. S. Sawhney and R. Kumar, "True multi-image alignment and its application to mosaicing and lens distortion correction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 3, pp. 235–243, Mar. 1999.

[16] M. Brown and D. G. Lowe, "Automatic panoramic image stitching using invariant features," *Int. J. Comput. Vis.*, vol. 74, no. 1, pp. 59–73, 2007.

[17] R. L. Marks, S. M. Rock, and M. J. Lee, "Real-time video mosaicking of the ocean floor," *IEEE J. Ocean. Eng.*, vol. 20, no. 3, pp. 229–241, Jul. 1995.

[18] N. Gracias and J. Santos-Victor, "Underwater video mosaics as visual navigation maps," *Comput. Vis. Image Understanding*, vol. 79, no. 1, pp. 66–91, 2000.

[19] Y. Rzhanov, L. M. Linnett, and R. Forbes, "Underwater video mosaicing for seabed mapping," in *Proc. Int. Conf. Image Process.*, 2000, vol. 1, pp. 224–227.

[20] O. Pizarro and H. Singh, "Toward large-area mosaicing for underwater scientific applications," *IEEE J. Ocean. Eng.*, vol. 28, no. 4, pp. 651–672, Oct. 2003.
[21] A. Elibol, N. Gracias, and R. Garcia, "Fast topology estimation for image mosaicing using adaptive information thresholding," *Robot. Autonom. Syst.*, vol. 61, no. 2, pp. 125–136, 2013.
[22] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[23] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, 1999, vol. 2, pp. 1150–1157.
[24] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.
[25] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.
[26] P. J. Rousseeuw, "Least median of squares regression," *J. Amer. Stat. Assoc.*, vol. 79, no. 388, pp. 871–880, 1984.
[27] N. Gracias and J. Santos-Victor, "Underwater mosaicing and trajectory reconstruction using global alignment," in *Proc. MTS/IEEE Conf. Exhib. OCEANS*, 2001, vol. 4, pp. 2557–2563.
[28] J. Ferrer, A. Elibol, O. Delaunoy, N. Gracias, and R. Garcia, "Large-area photo-mosaics using global alignment and navigation data," in *Proc. MTS/IEEE OCEANS Conf.*, Vancouver, BC, Canada, 2007, pp. 1–9.
[29] R. Prados, R. Garcia, N. Gracias, J. Escartin, and L. Neumann, "A novel blending technique for underwater gigamosaicing," *IEEE J. Ocean. Eng.*, vol. 37, no. 4, pp. 626–644, Oct. 2012.
[30] M. Uyttendaele, A. Eden, and R. Skeliski, "Eliminating ghosting and exposure artifacts in image mosaics," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2001, vol. 2, pp. 509–516.
[31] A. Agarwala *et al.*, "Interactive digital photomontage," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 294–302, 2004.
[32] P. J. Burt and E. H. Adelson, "A multiresolution spline with application to image mosaics," *ACM Trans. Graph.*, vol. 2, no. 4, pp. 217–236, 1983.
[33] C.-T. Hsu and J.-L. Wu, "Multiresolution mosaic," *IEEE Trans. Consum. Electron.*, vol. 42, no. 4, pp. 981–990, Nov. 1996.
[34] A. Agarwala, "Efficient gradient-domain compositing using quadtrees," *ACM Trans. Graph.*, vol. 26, no. 3, 2007, Art. no. 94.
[35] J. Escartín *et al.*, "Globally aligned photomosaic of the lucky strike hydrothermal vent field (mid-Atlantic ridge, 37 18.5 n): Release of georeferenced data, mosaic construction, and viewing software," *Geochem., Geophys., Geosyst.*, vol. 9, no. 12, 2008.
[36] D. Lirman *et al.*, "Damage and recovery assessment of vessel grounding injuries on coral reef habitats by use of georeferenced landscape video mosaics," *Limnol. Oceanogr., Methods*, vol. 8, no. 3, pp. 88–97, 2010.
[37] L. J. Dagum, F. J. Corpuz, M. Soriano, J. Jauod, E. Capili, and R. J. Judilla, "Stitching algorithm applied to camera array images for the visualization of Tubbataha reef grounding scars," in *Proc. OCEANS Conf.*, 2014, pp. 1–7.
[38] S. Rende *et al.*, "Advances in micro-cartography: A two-dimensional photo mosaicing technique for seagrass monitoring," *Estuarine, Coastal Shelf Sci.*, vol. 167, Part B, pp. 475–486, 2015.
[39] H. Singh, J. Howland, and O. Pizarro, "Advances in large-area photomosaicking underwater," *IEEE J. Ocean. Eng.*, vol. 29, no. 3, pp. 872–886, Jul. 2004.
[40] O. Pizarro, R. M. Eustice, and H. Singh, "Large area 3-d reconstructions from underwater optical surveys," *IEEE J. Ocean. Eng.*, vol. 34, no. 2, pp. 150–169, Apr. 2009.
[41] N. Gracias and J. Santos-Victor, "Automatic mosaic creation of the ocean floor," in *Proc. OCEANS Conf.*, 1998, vol. 1, pp. 257–262.
[42] Y. Rzhanov, L. Mayer, S. Beaulieu, T. Shank, S. Soule, and D. Fornari, "Deep-sea geo-referenced video mosaics," in *Proc. OCEANS Conf.*, 2006, pp. 1–6.
[43] F. Bellavia *et al.*, "Piecewise planar underwater mosaicing," in *Proc. OCEANS Conf.*, 2015, pp. 1–7.
[44] A. Elibol, J. Kim, N. Gracias, and R. Garcia, "Fast underwater image mosaicing through submapping," *J. Intell. Robot. Syst.*, vol. 85, pp. 167–187, 2016.
[45] "Autostitch: A new dimension in automatic image stitching." [Online]. Available: http://matthewalunbrown.com/autostitch/autostitch.html, Accessed on: Mar. 29, 2018.
[46] "Microsoft image composite editor." [Online]. Available: http://research.microsoft.com/en-us/um/redmond/projects/ice/, Accessed on: Mar. 29, 2018.
[47] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
[48] G. Baugh, A. Kokaram, and F. Pitié, "Advanced video debanding," in *Proc. 11th Eur. Conf. Visual Media Prod.*, 2014, Paper 7.
[49] X. Jin, S. Goto, and K. N. Ngan, "Composite model-based dc dithering for suppressing contour artifacts in decompressed video," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2110–2121, Aug. 2011.
[50] The Mathworks, Inc., "Matlab (r2015a)." [Online]. Available: https://uk.mathworks.com/products/matlab/, Accessed on: Mar. 29, 2018.
[51] A. Vedaldi, "SIFT for Matlab," 2006. [Online]. Available: http://vision.ucla.edu/~vedaldi/code/sift.html, Accessed on: Mar. 29, 2018.
[52] M. S. Hitam, E. A. Awalludin, W. N. J. H. W. Yussof, and Z. Bachok, "Mixture contrast limited adaptive histogram equalization for underwater image enhancement," in *Proc. Int. Conf. Comput. Appl. Technol.*, Jan. 2013, pp. 1–5.
[53] J. Y. Chiang and Y. C. Chen, "Underwater image enhancement by wavelength compensation and dehazing," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1756–1769, Apr. 2012.
[54] S. Serikawa and H. Lu, "Underwater image dehazing using joint trilateral filter," *Comput. Electr. Eng.*, vol. 40, no. 1, pp. 41–50, 2014.
[55] K. Sooknanan, A. Kokaram, D. Corrigan, G. Baugh, J. Wilson, and N. Harte, "Improving underwater visibility using vignetting correction," in *Proc. SPIE Visual Inf. Process. Commun.*, Burlingame, CA, USA, Jan. 2012, Paper 83050M.
[56] P. List, A. Joch, J. Lainema, G. Bjontegaard, and M. Karczewicz, "Adaptive deblocking filter," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 614–619, Jul. 2003.

David Corrigan received the B.A./B.A.I. degree in computer and electronic engineering and the Ph.D. degree under the supervision of Prof. A. Kokaram from Trinity College Dublin, Dublin, Ireland, in 2003 and 2008, respectively. The subject of his doctoral thesis was on the relationship between motion estimation reliability and archived film restoration.

Since then, he has been working in the research team with The Foundry, a leading provider of visual-effect software to the cinema postproduction industry, and has spent three years as a Research Fellow with the Sigmedia Group, Trinity College Dublin. From 2011 to 2016, he was an Assistant Professor at Trinity College Dublin where his research interests were in applications of video processing and computer vision in marine surveillance, life sciences, and the creative industries. He is currently a Principal Engineer with the Video Competence Centre, Huawei Irish Research Centre, Dublin.

Ken Sooknanan was born in San Juan, Trinidad, in 1980. He received the M.Sc. degree in digital systems from the University of the West Indies, Trinidad, Jamaica, in 2010 and the Ph.D. degree in electronic and electrical engineering from Trinity College Dublin, Dublin, Ireland, in 2015.

Since 2015, he has been working at the Department of Information and Communication Technology, University of Trinidad and Tobago, Trinidad, as an Assistant Professor. His research interests include image processing and machine learning for marine and medical applications.

Jennifer Doyle received the B.Sc. degree in biology from the National University of Ireland (NUI), Maynooth, Ireland, and the Diploma in fisheries management from NUI, Cork, Ireland, in 1996 and 1997 respectively.

She works as a Scientific and Technical Officer at the Marine Institute, Galway, Ireland. Since 2002, she has been involved with underwater TV surveys used to assess Nephrops stocks of commercial interest to the Irish fishing industry. She has also been the Chair of the International Council for the Exploration of the Sea Workshop on Nephrops Burrow Counting in 2016.

Colm Lordan was born in Cork, Ireland, in 1971. He received the B.Sc. (honors) degree in applied ecology and the Ph.D. degree in the field of fisheries science from the University of Cork, Cork, Ireland, in 1993 and 2001, respectively.

Since 1998, he has been working at the Marine Institute, Galway, Ireland. He leads a team of scientist conducting fisheries surveys, stock assessment, and providing scientific advice on demersal fish and Nephrops stocks. He has been developing and expanding the MI UWTV Survey Programme for Nephrops since 2002. He chaired the International Council for the Exploration of the Seas Working Groups on Nephrops Surveys in 2007 and 2012–2014.

Anil Kokaram received the Ph.D. degree in signal processing from the Department of Engineering, Cambridge University, Cambridge, U.K., in 1993.

He is currently a Professor and also the Head of the Department of Electronic and Electrical Engineering, Trinity College Dublin, Dublin, Ireland. From 2011 to 2017, he was the Lead of the Media Algorithms Team at YouTube/Google. In 1998, he founded a media DSP group at Trinity www.sigmedia.tv. His expertise is in the broad areas of DSP for video processing, Bayesian inference, and motion picture engineering. He has published more than 100 refereed papers in these areas. He was the founder of a company (GreenParrotPictures) producing video enhancement software that was acquired by Google in 2011.

Dr. Kokaram was a recipient of the Science and Engineering Academy Award from the American Academy of Motion Picture Arts and Sciences for his work in video processing for postproduction applications.

He was a former Associate Editor for the IEEE TRANSACTIONS ON VIDEO TECHNOLOGY and the IEEE TRANSACTIONS ON IMAGE PROCESSING.